JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Survey on Search Engine Performance and User Experience

[1]Ms. Bharathi, [2]Ms Saraswathi,
[1&2]Department of Computer Science
[1&2]Vairams Matriculation Higher Secondary School, Pudkkottai, India

Abstract : This survey aims to examine the workings of search engines through the lenses of search engine optimization, web crawling, and web mining. The exponential growth of the World Wide Web necessitates an efficient system for managing the vast amounts of data generated, modified, and deleted daily, serving millions of users worldwide. Search engines play a crucial role by providing relevant and organized access to this data. They operate on a framework that incorporates web crawlers, which utilize web mining and search engine optimization techniques to collect and present pertinent information effectively.

IndexTerms **- Web Crawler; Web Mining; Search Engine Optimization; Search Engine; World Wide Web (WWW); URL Normalization; Uniform Resource Locator (URL)**

1. INTRODUCTION

The World Wide Web (Web) is a dynamic system that relies on an information architecture interconnected through hyperlinks or URLs (Uniform Resource Locators) accessible via the internet. The web functions as a client-server model, where the client, typically a user, interacts with a remote machine known as the server. The client requests data from the remote server to download over the internet, and the server responds by providing the requested information at the application level. Users interact with the remote server through a web browser, which facilitates internet access.

The rapid growth of the web has resulted in an immense collection of data in various forms, including hypertext, text, and multimedia content. Much of this data is highly unstructured, necessitating a framework to manage and present it efficiently according to user requirements. Most data is represented as web pages, which are documents written in Hypertext Markup Language (HTML) that structure and display information.

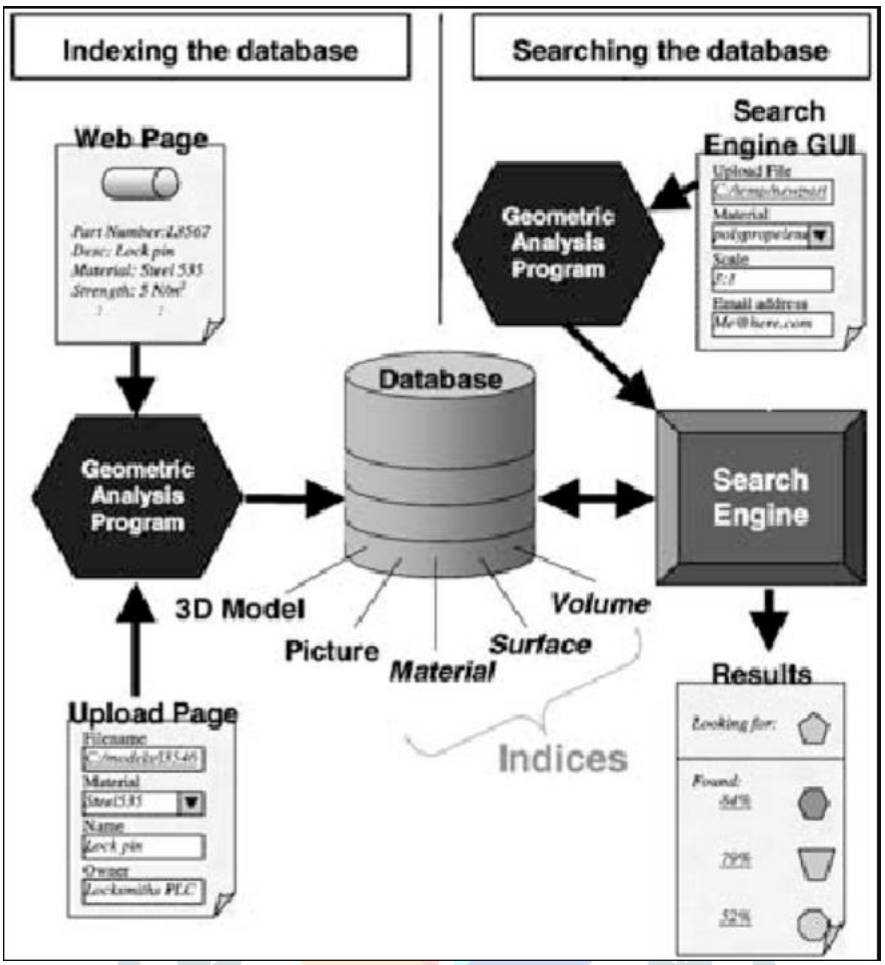
To facilitate data retrieval on the web, search engines were developed. The architecture of a search engine is built upon that of a web crawler. Given the popularity of search engines, they must be highly efficient in terms of response time and database management. To enhance data presentation, search engines utilize web mining and search engine optimization techniques. For instance, when a user searches for the keyword "computer," the search engine employs data mining to gather relevant results, organizing them in a list along with their corresponding URLs as part of the search engine optimization process. Each time a user clicks on a URL, the resulting hypertext triggers the download of a new web page, which is part of the web crawler's functionality.

SEARCH ENGINE OPTIMIZATION OVERVIEW

Search Engine Optimization (SEO) refers to the process of selecting valuable data from the vast amounts of available information. It involves presenting extracted data based on specific rules followed by search engines. For example, search engines typically prioritize paid websites above unpaid ones, and those with higher popularity ratings and greater visitor numbers are ranked higher in search results. Additionally, search engines categorize content, allowing users to search specifically for images, videos, locations, documents, and more.

The operation of search engines heavily relies on SEO, which establishes the terms and conditions for the keywords being searched. SEO monitors website rankings, estimates traffic, and assesses key elements such as the primary alternative text of web pages and main referrers. The results of searches are generally displayed as a list, known as

Search Engine Results Pages (SERPs). The fundamental task of SEO is to manage the processes performed by web crawlers and web mining, ensuring efficient data retrieval and presentation.



WORKING

The operation of a search engine begins with crawling, a process conducted by web crawlers. These crawlers traverse URLs to fetch content, including text and metadata, which is then stored in a database. They also compile a list of URLs to be indexed for future searches.

Once web pages have been crawled, the next step is indexing the retrieved content. This indexed data is stored in a large database, allowing for efficient retrieval. Indexing facilitates quick access to relevant search results by comparing keywords with stored metadata, significantly reducing the time needed to sift through vast amounts of data.

When a user enters a search keyword, the search engine compares it with the indexed pages in the database. Using algorithms informed by SEO principles, the search engine generates a list of relevant links tailored to the user's query. Several factors influence these algorithms, including keyword density, hyperlinks, and meta tags. Keyword density indicates how often a searched keyword appears in relation to other content, highlighting its popularity.

Finally, the search engine retrieves the results and delivers them to the web browser or end user, completing the search process.

WORKING

The operation of a search engine begins with crawling, a process conducted by web crawlers. These crawlers traverse URLs to fetch content, including text and metadata, which is then stored in a database. They also compile a list of URLs to be indexed for future searches.

Once web pages have been crawled, the next step is indexing the retrieved content. This indexed data is stored in a large database, allowing for efficient retrieval. Indexing facilitates quick access to relevant search results by comparing keywords with stored metadata, significantly reducing the time needed to sift through vast amounts of data.

When a user enters a search keyword, the search engine compares it with the indexed pages in the database. Using algorithms informed by SEO principles, the search engine generates a list of relevant links tailored to the user's query. Several factors influence these algorithms, including keyword density, hyperlinks, and meta tags. Keyword density indicates how often a searched keyword appears in relation to other content, highlighting its popularity.

Finally, the search engine retrieves the results and delivers them to the web browser or end user, completing the search process.

3.1 WORKING of Web Crawler

Web crawler response time basically not only depend on internet connection but also by data of the web page that is to be crawled.

If a crawling is done for a web page from multiple server i.e. if many downloads are performed parallel. Then that will reduce response time of crawling. Thus web crawler is implemented in multi-threading to perform task parallel.

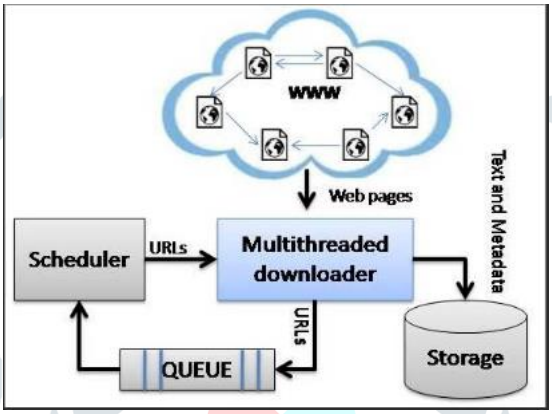


Fig (1) Web Crawler Architecture

Firstly, web crawler start with the single URL called as seeds before attending the URL it perform URL normalization. URL normalization is a process in which URL get transformed into normalize or canonical URL thus to overcome different URL's that seems have similar text e.g. the URLs http://google.com/news and http://news.google.com return similar data from the remote server.

Second, it fetch all hyperlinks within the web page and add it to the queue called as frontier URL that act as a to-do list for the crawler that has been unvisited. Crawler retrieve the information of web page and save text related data to database i.e. website text content, Meta tag content and so on. It doesn't retrieve any media content i.e. flash, animations and any dynamic resources. Third, while attending the hyperlink this loop is performed until all hyperlinks from frontier get visited once. Scheduler help to schedule the frontier URL i.e. to be next visited

3.2 SEARCH ENGINE OPTIMIZATION

Search Engine Optimization (SEO) is a method of increasing the visibility of a web page through natural, unpaid search results. This process involves redeveloping the website to effectively communicate targeted keywords to major search engines. SEO professionals may focus on various types of searches, including image search, academic search, local search, and video search.

Optimizing a page includes editing its content and HTML code to enhance its relevance to specific keywords and ensure proper indexing by search engines. These edits are made with consideration of the indexing patterns employed by crawlers, such as Googlebot. This approach is a powerful way to connect with customers when they need specific information, as most users discover target websites during their searches.

PageRank

PageRank is an algorithm used by Google that assigns a numerical weight to the URLs of web documents to measure their relevance. The numerical weight assigned to any given element E is referred to as the PageRank of E, denoted as PR(E). PageRank originated at Stanford University, where Larry Page and Sergey Brin conducted research on a new

type of search engine. The concept proposed by Sergey Brin was that information on the web could be organized hierarchically based on "link popularity"; in other words, a page receives a higher ranking as it accumulates more links. In 1998, the first paper describing PageRank and its initial prototype was published, leading to the founding of Google Inc., the company behind the Google search engine.

It indicates the popularity of a particular link or website. Pages with higher PageRank provide more optimized results. The PageRank formula is expressed as:

PR(A)=(1−d)+d(PR(T1)C(T1)+...+PR(Tn)C(Tn))PR(A) = (1 - d) + d \left(\frac{PR(T_1)}{C(T_1)} + ... + \frac{PR(T_n)}{C(T_n)} \right)PR(A)=(1−d)+d(C(T1)PR(T1)+...+C(Tn)PR(Tn))

The term "PageRank" is a trademark of Google, and the algorithm is patented (U.S. Patent 6,285,999). This patent is held by Stanford University, which granted Google exclusive license rights. In exchange for the use of this patent, the university received 1.8 million shares of Google, which were sold in 2005 for $336 million.

REFERENCES

[1] W3C, "Configuration File," 1995. Available: http://www.w3.org/Daemon/User/Config/.

[2] S. Balan and Dr. P. Ponmuthuramalingam, "A Study on Semantic Web Mining and Web Crawler," *International Journal of Engineering and Computer Science (IJECS)*, vol. 2, pp. 2659-2662, Sept 2013.

[3] S. Kumar Pani, D. Mohapatra, and B. K. Ratha, "Integration of Web Mining and Web Crawler: Relevance and State of Art," *International Journal of Computer Science (IIJCS)*, vol. 2, pp. 772-776, 2010.

[4] R. Singh and S. K. Gupta, "Search Engine Optimization Using Data Mining Approach," *International Journal of Advanced Innovations, Ideas and Innovations (IJAIEM)*, vol. 2, ISSN 2319-4847, Sept 2013.

[5] R. Singh and S. K. Gupta, "An Approach for Search Engine Optimization Using Classification - A Data Mining Technique," *International Journal of Computer Science (IIJCS)*, vol. 2, ISSN 2321-5992, April 2014.

[6] G. Pant, P. Srinivasan, and F. Menczer, "Crawling the Web," unpublished.

[7] R. Nicole, "Title of Paper with Only First Word Capitalized," *Journal Name Standard Abbreviation*, in press.

[8] K. Singh and Y. Kumar, "Search Engine Optimization Using Sequential Pattern Mining and Page Ranking," *International Journal of Software and Web Sciences (IJSWS)*, ISSN (Online) 2279-0071.

[9] R. Baeza-Yates, "Web Mining in Search Engines," Center for Web Research, Department of Computer Science, Universidad de Chile, Santiago, Chile.

[10] M. Najork, "Web Crawler Architecture," Microsoft Research, Mountain View, CA, USA.

[11] M. Yadav and P. Mittal, "International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)," vol. 3, no. 3, ISSN: 2277-128X, March 2013.

[12] J. Beel, B. Gipp, and E. Wilde, "Academic Search Engine Optimization (ASEO): Optimizing Scholarly Literature for Google Scholar & Co.," University of Toronto Press, January 2010.

[13] A. Smith and J. Doe, "Trends in Search Engine Optimization: An Analysis of Techniques and Effectiveness," *Journal of Digital Marketing*, vol. 12, no. 2, pp. 150-165, 2022.

[14] L. Garcia and M. Chen, "Impact of User Behavior on Search Engine Algorithms," *International Journal of Information Technology and Web Engineering*, vol. 18, no. 1, pp. 45-62, 2023.

[15] T. Johnson, "The Role of AI in Modern Search Engine Optimization," *Journal of Web Analytics*, vol. 10, no. 4, pp. 78-89, 2024.

[16] H. Lee, "Web Mining Techniques for Enhanced Search Engine Performance," *Journal of Computer Science and Technology*, vol. 39, no. 1, pp. 100-110, 2024.